Industry and R&D Needs: Defogging Key Issues in M&V2.0

Jessica Granderson

Team:
Samuel Fernandes, Samir Touzani
Lawrence Berkeley National Laboratory

David Jump
KW Engineering

BERKELEY LAB
Lawrence Berkeley National Laboratory

Outline

- Intro to M&V2.0 and webinar focus
- Who is doing what on the general topic
- What do we know about commercial M&V2.0, what is next, and where are we going?

What is M&V2.0?

- Generally understood as: use of more data data (interval or volume), analytics, computation at scale
  - to streamline the M&V process through semi/automation

- Delivered in proprietary tools, 'open' algorithms

What are the potential benefits of M&V2.0?

- Increase visibility, quickly obtain ongoing and interim results feedback
  - Increase savings and enhance customer experience?

- Automate parts of the process that computers do well, streamline data acquisition and processing
  - Reduce time and cost?
  - Increase throughput, number of projects going through the pipeline?

What is the vision for where we might end up?

- New M&V2.0 methods can be objectively tested as industry continues to innovate and new data source become available
- Multiple real-world pilots are used to assess M&V 2.0 value proposition
  - Cost, accuracy, time, tradeoffs vs. traditional M&V
  - Value of continuous feedback in increasing savings as well as customer value and experience – for both residential as well as commercial
- Processes/work flows are established to leverage automation while using engineering expertise where needed to maintain a quality result
- Analytical solutions to flag the non-routine adjustments are developed and tested for effectiveness
- Industry establishes acceptable levels of uncertainty and confidence, and documentation requirements for transparent evaluation

What is new about M&V2.0? What is not new?

- M&V2.0 tools are built upon savings estimation techniques that have been used for decades
  - Comparison group analyses,
  - IPMVP Options B&C, whole-building and submeter-based
  - IPMVP Option D, calibrated simulation modeling

- What's new is:
  - Degree of automation in data acquisition, and model creation
  - Granularity and volume of data can improve quality of result
  - Potential for continuous feedback
  - Integration of M&V capability with other analyses for operational efficiency
    - eg load visualization, portfolio tracking, end-use monitoring, etc.
  - Software as a service offerings for owners, managers, program administrators

Two examples

ENERGYSAVVY

- Customer engagement
- Program administrators
- Continuous savings feedback
- Net savings, comparison group billing analysis
- Residential

lucid

- Operational efficiency, SEM, MBCx
- Owners and operators
- Continuous savings feedback
- Gross savings, pre/post whole-building or submeter Option B or C
- Commercial

Screen shots of M&V2.0 capability

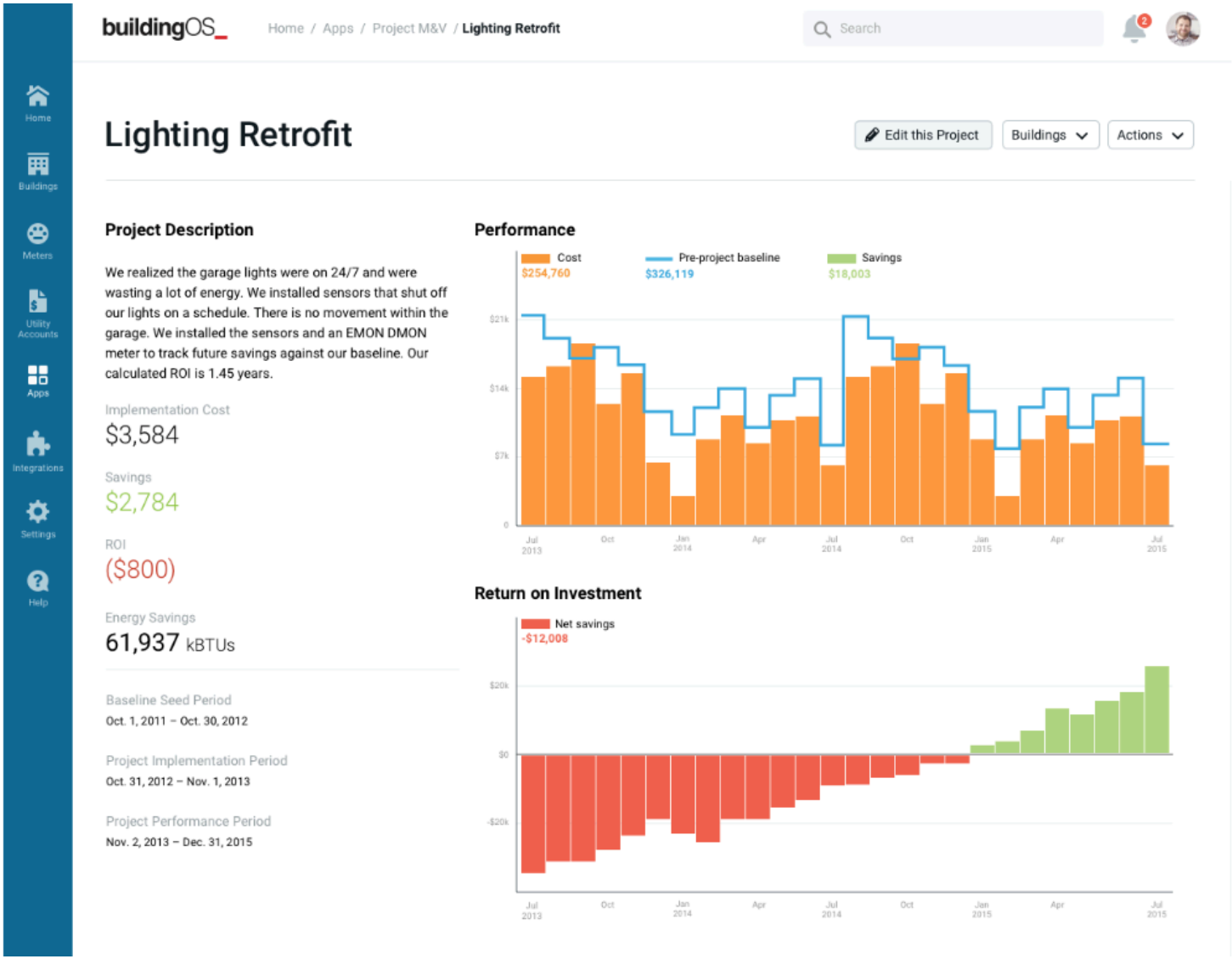


Image Source: Lucid

BERKELEY LAB
Lawrence Berkeley National Laboratory

Screen shots of M&V 2.0 capability

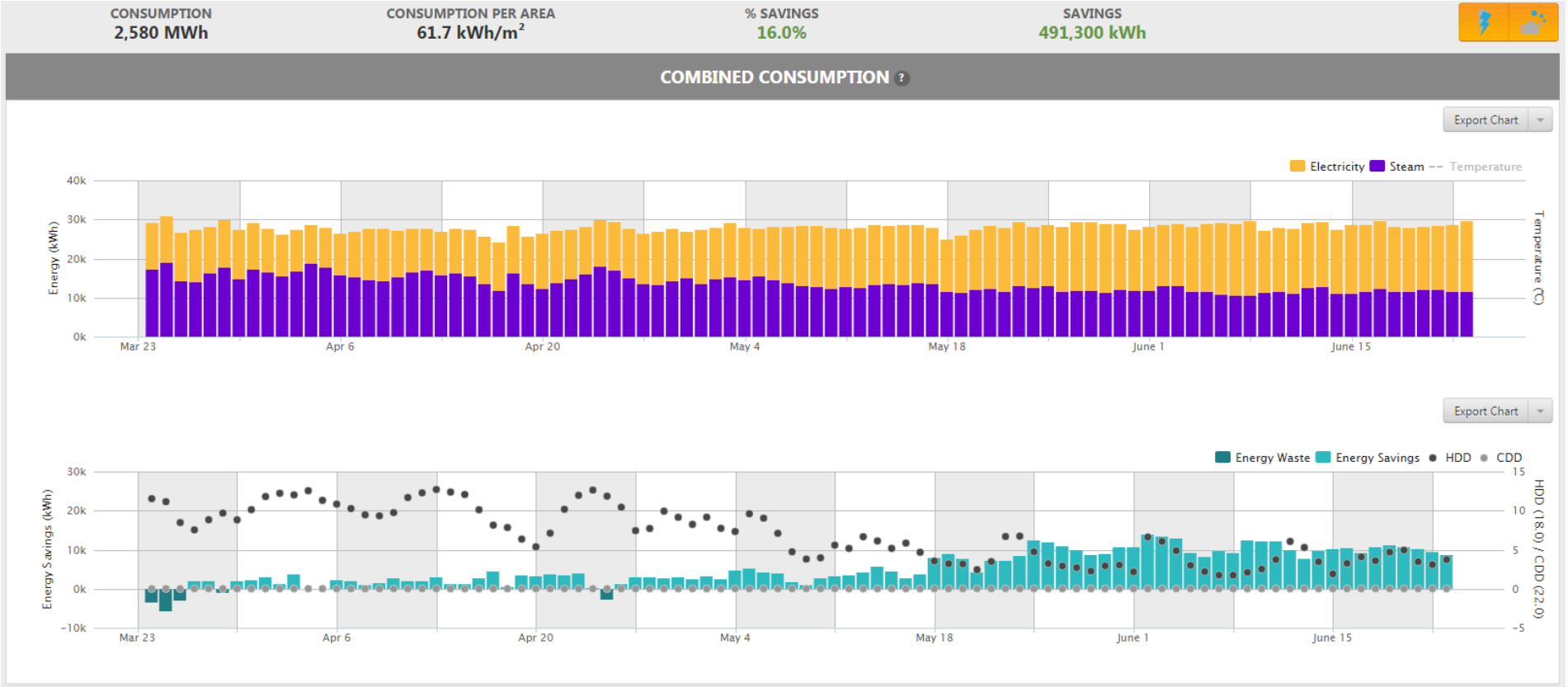


Image Source: EnerNOC

BERKELEY LAB
Lawrence Berkeley National Laboratory

Screen shots of M&V2.0 capability

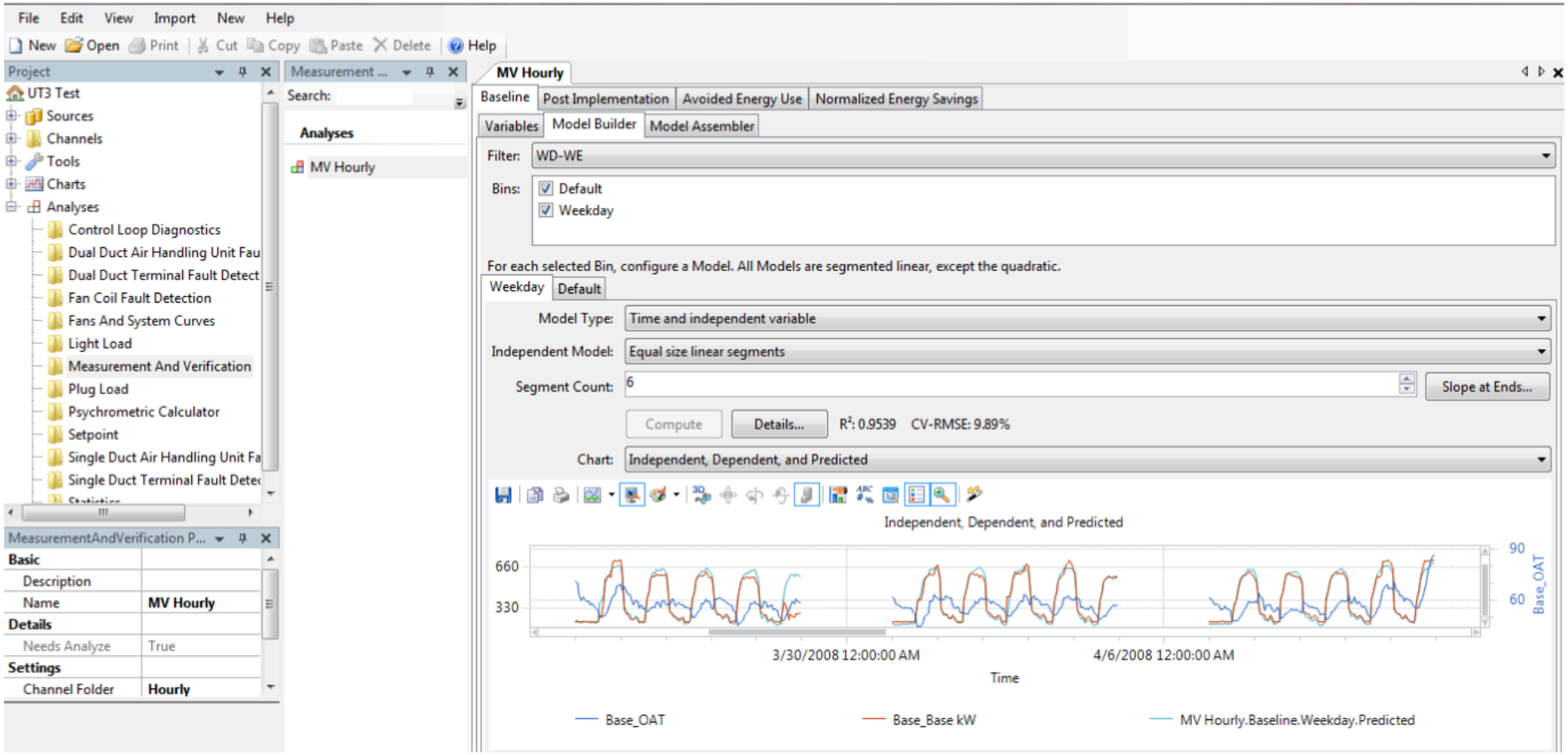


Image Source: Universal Translator 3

A diversity of savings estimation approaches is used today

Approach	Meter based?	Net or Gross?	Program/measure sweet spot
Deemed values	Not directly	Gross, Net in some cases	Efficient equipment replacement/installation
Engineering estimates, calculated	Not typically	Gross, Net	Custom industrial and large commercial; new construction
Billing Analysis that can include comparison groups, randomized control trial, or quasi-experimental	Yes, with other data	Usually Net	Programs w large numbers: residential, behavioral, small savings/site
Calibrated simulation modeling (IPMVP Option D)	No (except the calibration)	Gross, Net	Retrofit, large commercial
***Whole-building and retrofit isolation M&V (IPMVP Option C, B)**	**Yes, with other data**	**Gross**	**Commercial, multi-measure, interactive effects, operational measures**
…	…	…	…

***This is the focus of the today's conversation**

Where are meter-based approaches most appropriate?

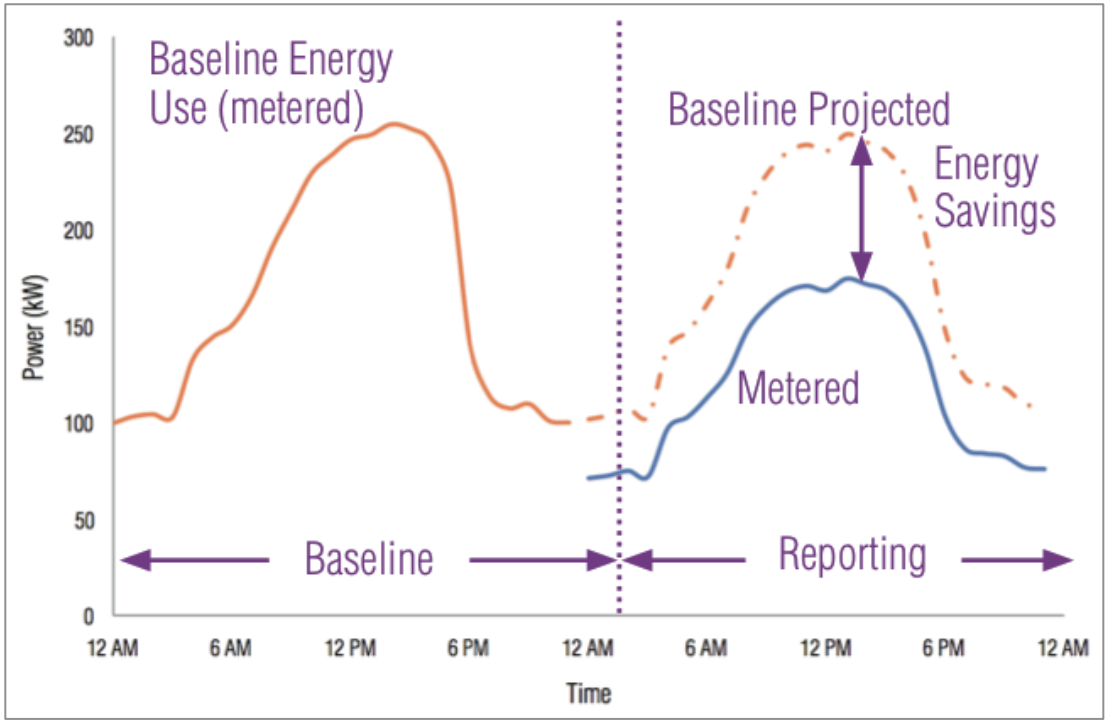
- 'Predictable' buildings
  - Weather sensitive, regularly scheduled
- Projects with multiple and interactive measures
  - Affecting several building systems (HVAC, lighting, etc.)
- Difficult to quantify measures

  Duct sealing, envelope upgrades, etc.
- Projects with larger savings, 'above the noise'
- **Measures using existing condition as baseline**
  - **Retrocommissioning, behavioral, operational**

Promising opportunities associated with meter-based M&V approaches

- Enabling delivery of whole-building programs that combine strategies for deep savings
- Enabling pay-for-performance programs
- Scalability and streamlining
  - Reduce labor time and costs
  - Maintain an accurate result
  - Quickly obtain ongoing and interim results
  - Increase throughput, number of projects

How are meter-based site savings quantified?



Metering at whole-building (Option C), or submetered measure isolation level (Option B)

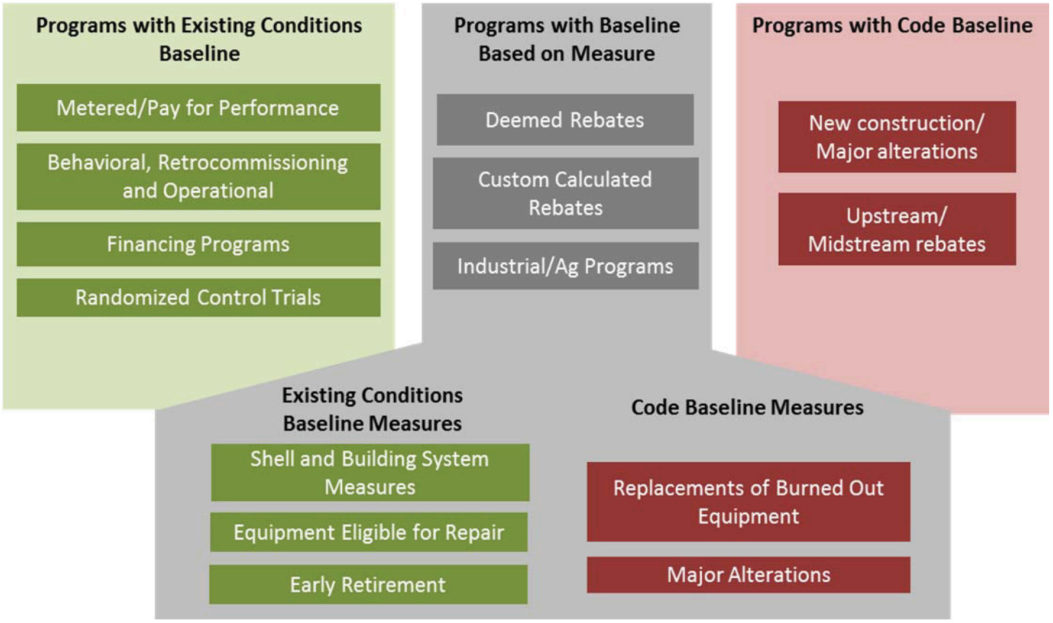
In M&V2.0 tools baselines are automatically created with meter and weather data feeds

User enters date of measure implementation, savings are calculated by the tool

Relevant California Activities

- CA AB802: CPUC to authorize programs using "normalized meter-based" energy savings (existing conditions baseline) for:
  - To- and beyond-code savings, and retrocommissioning, operational, behavioral programs
  - Counting savings towards goals when feasible and cost-effective
- CPUC providing guidance on where existing use baselines are/ not appropriate, EM&V plans

Figure 3: Proposed Baseline Framework



Other Relevant Activities

- RMI e-Lab Accelerator cross-stakeholder group group, more detailed articulation of M&V2.0, potential benefits, outstanding industry needs
- CEE Guidebook resource to understand uncertainty principles for whole-building M&V approaches, in context of whole-building program deign and delivery
- EVO has started an M&V2.0 group to determine how IMPVP will address the topic
- ASHRAE technical committees discussing 'standard methods of test'

Break to participant poll questions, report out in real-time to the group as results come in.

1. Are you involved in, or going to be involved in any programs that rely on an existing use baseline? [Y/N/maybe]

2. Are you interested in exploring 2.0 tools and methods in your work? [Y/N/maybe]

Motivating Industry Questions, R&D Approach, and Highlights

Industry questions motivate LBNL's R&D

- Are these proprietary tools reliable?
- How can I verify their accuracy and compare them?
- Are proprietary tools any better or worse than standard regressions?
- Even if a tool is generally robust, how do I know that it will work for my specific projects or program?
- How "big" do my savings have to be to use these approaches?
- How do I know that a robust tool was applied to generate a quality result?

Four-step R&D approach to answer these questions

1. **Population-level (many buildings) M&V2.0 testing to verify general, overall robustness, compare and contrast tools**

2. **'Off-line' demonstration of promising models with historic utility program data**

3. Identification of reporting requirements and quantitative acceptance criteria for savings claims (in progress)

4. Larger pilots, demonstrations on 'live' programs (future)

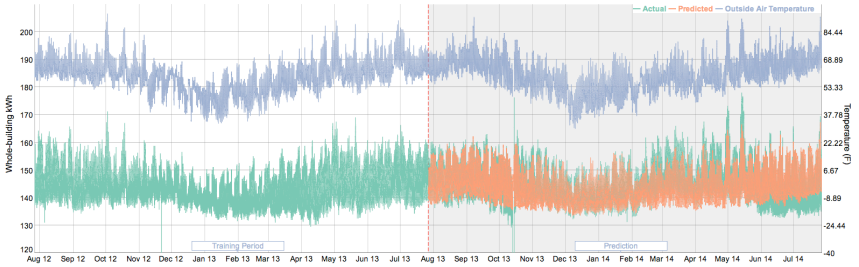
1. Population-level general testing and tool comparison

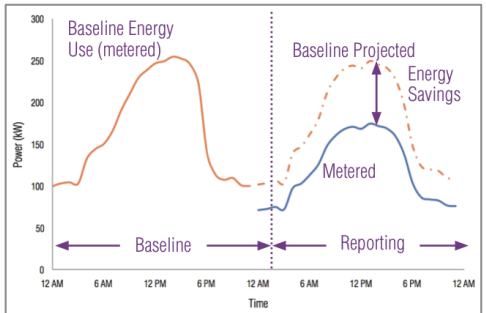
- Tested accuracy of baseline projections in proprietary tools and open standard models against data set from 500-600 untreated buildings

- Given 12mo whole building interval data, predicted 12mo of energy use
  - Within {-4, 5}% error for a full half of the buildings, CV(RMSE) well within industry guidelines, errors even smaller when aggregating buildings into portfolio
  - No clear 'winner' across 10 models

- No attempts to refine models based on expertise, knowledge of buildings, additional variables
  - Floor of predictive accuracy

- Test procedure is published, was used by PG&E to prequalify tools for pilot, is available for use by others

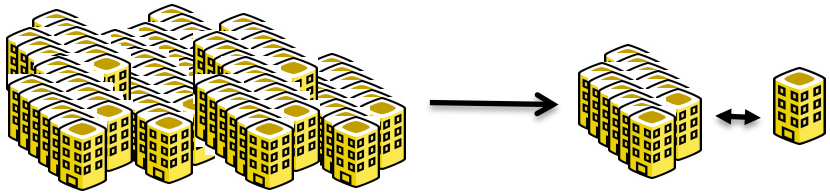




2. Demonstrating 2.0 tools with historic program data

- Given tools that generally predict energy well, use them to automatically quantify savings

- Develop practitioner workflows to leverage automation while retaining accuracy of the savings result
  - Many, but not all buildings are 'predictable'
  - Gross savings at the meter may not be gross savings due to the measure, i.e., non-routine adjustments may be needed
  - Use uncertainty analysis to quantify *accuracy of the savings results when applied to specific projects/buildings/programs*



We draw from ASHRAE Guideline 14

- Provides **recommendations for accuracy in M&V**
- Covers **'goodness of fit' between the model and the baseline period data**, with suggested thresholds for bias (NMBE) and CV(RMSE)
- Covers suggested formulae to quantify **uncertainty due to error in the baseline model**
- Suggests **that fractional uncertainty be no more than 50% with at least 68% confidence** (what will EE programs require?)

Model demonstration with historic program data

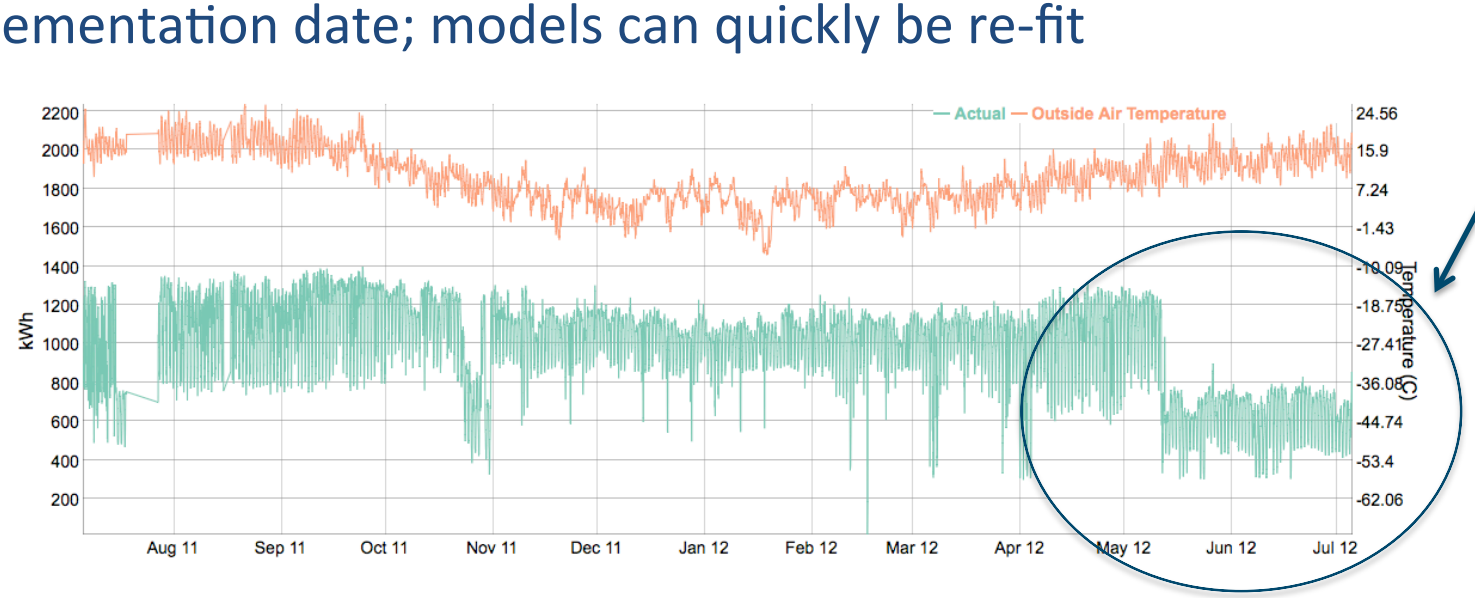
- Data from 51 buildings that underwent RCx and in some cases retrofits

- Preliminary workflow, drawing from ASHRAE Guideline 14
  - Auto fit the model to data from baseline period, and compute goodness of fit metrics
  - Set aside buildings that do not meet suggested fitness thresholds - these will require further investigation
  - For 'good' buildings auto compute savings and uncertainty using M&V 2.0 tool
  - Aggregate savings and uncertainties for each building to determine portfolio-level results

Findings from applying this workflow to historic program data

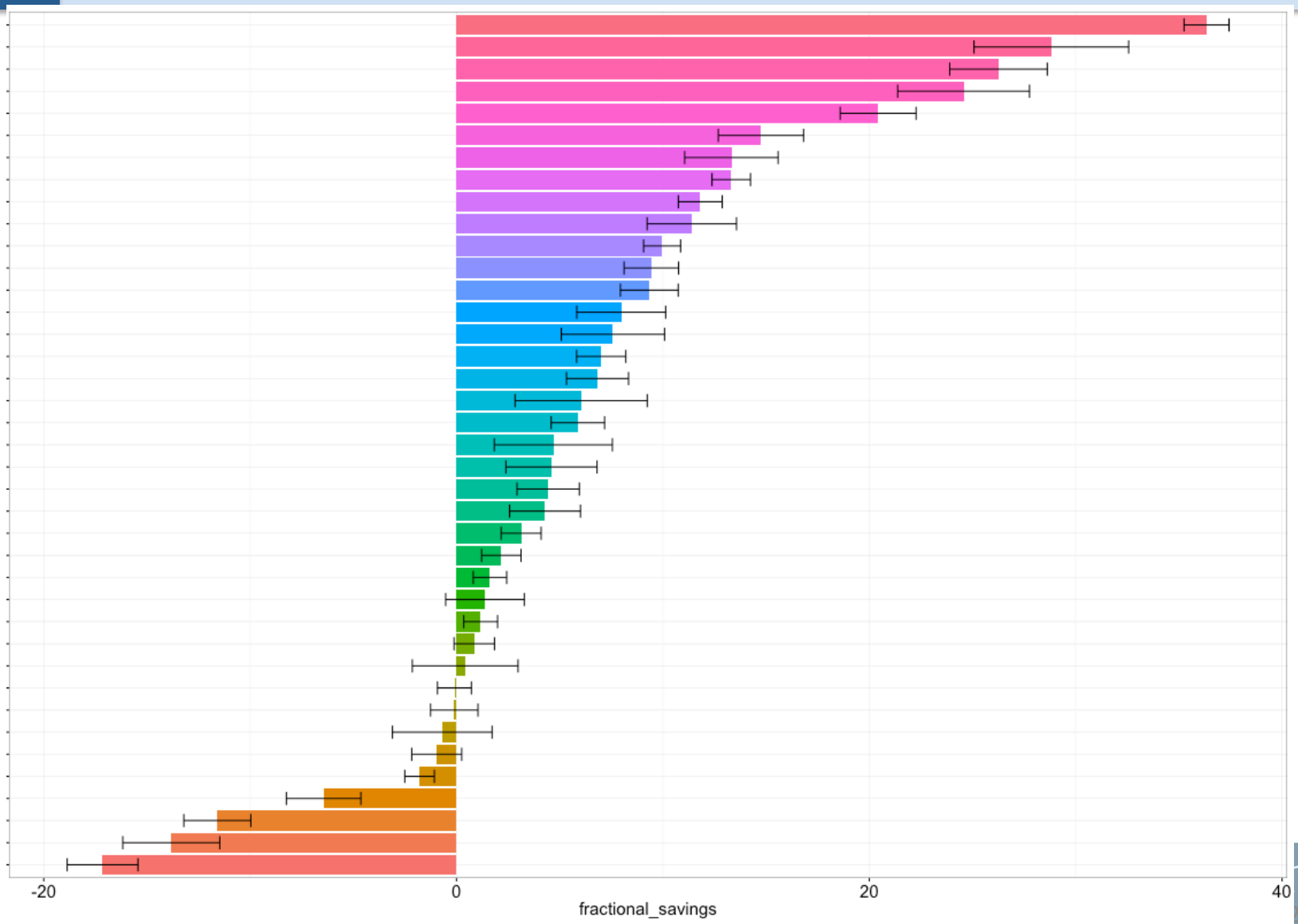
- Of the 51 buildings, 39 'passed' the goodness of fit tests using ASHRAE guidance

- Of the 12 that did not 'pass', 5 had incorrect documentation of measure implementation date; models can quickly be re-fit



- **For this data set, 44 of 51 buildings look to be well-suited to automated analysis; 7 may require more manual investigation**

Findings from uncertainty analysis with historic program data



Savings uncertainty ranges for each of 39 buildings, at 95% confidence level

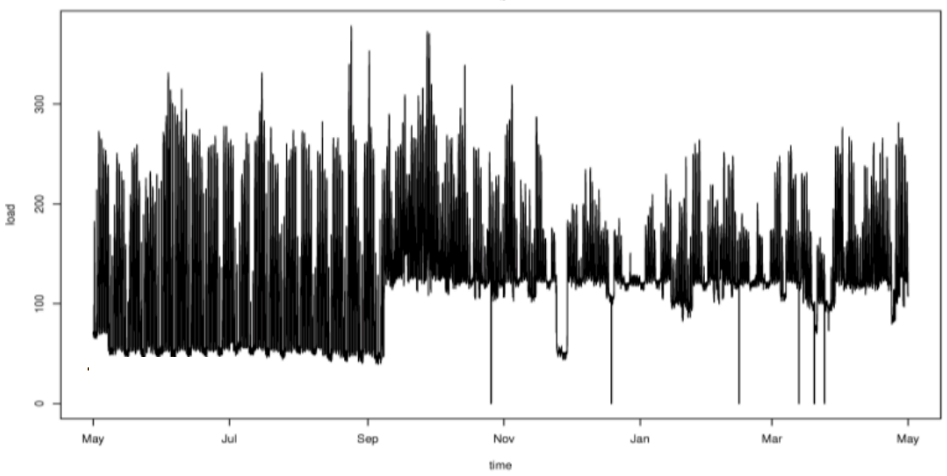
Summary of uncertainty findings in the demonstration on historic program data

- **32 of 39 individual buildings satisfied** or exceeded **ASHRAE** uncertainty requirements

- **At portfolio-level** for the aggregate of the 39 buildings, at 95% confidence level
  - Savings = 3.96% =/-.3, that is within confidence interval of [3.66%; 4.26%]
  - Aggregate ***far*** **exceeds ASHRAE guidance for sufficiency**

Some comments on non-routine adjustments

- Gross metered savings may not reflect gross program/measure savings
  - E.g. Occupancy or schedules may change or loads may be added/removed



- By definition, these Option-C compliant M&V2.0 baseline models do *not* handle NR Adj.

- It is possible that 2.0 analytics can flag cases where savings drop or increase unexpectedly, so that implementers can make timely inquiries of the site

Some comments on uncertainty, confidence, and documentation requirements

- General tool testing can tell us that we have good well-made hammers

- If we have well-made hammers, uncertainty and confidence can verify that we've driven our nails straight and true
  - But how straight do we need to be?
  - An how do we prove it to 3rd parties?
  - What documentation will we need?

80-20? 90-10? And how do we set these values?

Some comments on net, gross, other baselines and methods

- Even with deemed savings you commonly need to layer additional analysis to determine a net
- Existing conditions baselines are critical to less common programs that
  - Promise deep savings, offer opportunity beyond equipment-based measures
  - Focus on operational, retro-commissioning, behavioral, multi-measure, whole-building
- Calibrated simulation can be complex, costly and difficult to scale
- Comparison groups may not always be possible to establish for commercial buildings

Break to participant poll questions, report on results, and based on response take clarifying questions:

1. Do you see value in distinguishing between: a) general population level testing and tool comparison, and; b) assessment of tool accuracy for specific buildings, programs, projects? [Y/N/maybe]

Where Have We Gotten and Where are We Going?

Where have we gotten?

- Appreciation of the potential benefits of M&V2.0
- Replicable test procedures to assess overall robustness of M&V 2.0 tools for commercial buildings
  - *Many* models predict within a few percent for *many* buildings using commonly available data
  - Use by large utility to pre-vet vendors for pilot, published for ongoing use
- Initial exploration (ongoing beyond the 51 projects shown here)
  - High confidence and low uncertainty when applying M&V2.0 tools
  - Start on defining practitioner workflows to retain a quality result
  - Indication that with interval data savings may not have to be as big as 10% to 'see' at the whole-building level

Returning to the the vision for where we might end up

- **New M&V2.0 methods can be objectively tested** as industry continues to innovate and new data source become available
- Multiple **real-world pilots are used to assess M&V 2.0 value** proposition
  - Cost, accuracy, time, tradeoffs vs. traditional M&V
  - Value of continuous feedback in increasing savings as well as customer value and experience – for both residential as well as commercial
- **Processes/work flows are established to** leverage automation while using engineering expertise where needed to **maintain a quality result**
- Analytical **solutions to flag the non-routine adjustments** are developed and tested for effectiveness
- Industry establishes **acceptable levels of uncertainty and confidence, and documentation requirements** for transparent evaluation

What do we still need to know or do?

- Demonstrate 2.0 tools on more historic program data (ongoing)
- What do we do for buildings that don't have a good fit, and aren't well suited to the meter-based approach?
  - How can we leverage targeting and pre-screening
- How does M&V2.0 compare to traditional approaches, 'in the field'?
  - **Can we conduct a sufficient number of pilots and what 'proof' points should they be designed to produce?**
- How do we handle non-routine adjustments?
- What uncertainty, confidence, and documentation requirements are needed for evaluation?
- What group might serve as a testing body for new M&V2.0 tools to verify their general robustness?

Break to participant poll questions, report on results, and use results to seed coming discussion.

1. How important are pilots as a next step? [very, somewhat, not important]

2. What should *commercial* pilots aim to evaluate? [select all that apply: time and cost of M&V2.0 vs traditional M&V approaches; value of M&V2.0 in providing continuous feedback; uncertainty due to baseline error; how NR Adj are handled. Free response: other, please describe]

3. Are you interested in participating in *commercial* pilot design or implementation? [Y/N/maybe]

Break to participant poll questions, report on results, and use results to seed coming discussion.

1. Is tool testing important for the acceptance of M&V2.0 tools [Y/N]
2. Is uncertainty analysis important for the acceptance of these methods? [Y/N]
3. Are you interested in continuing to work on topics of tool testing and acceptance criteria? [Y/N/Maybe]

Questions and Discussion

Thank You!

For more information please contact Jessica Granderson
JGranderson@lbl.gov, 510.486.6792

For more detailed reports and presentations: eis.lbl.gov